ADA030994

DECISION MAKING RESEARCH

DEPARTMENT OF PSYCHOLOGY
DEPARTMENT OF MANAGEMENT AND ORGANIZATION
UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON

DECISION MAKING RESEARCH

DEPARTMENT OF PSYCHOLOGY

UNIVERSITY OF WASHINGTON

SEATTLE, WASHINGTON

UTILITY: RECENT THEORY AND SOME APPLICATIONS

Marcia D. Deaton

University of Washington
Seattle, Washington

Technical Report 76-4

September 1976

Office of Naval Research Contract N00014-76-C-0193

(Terence R. Mitchell and Lee Roy Beach, Investigators)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

14

1. REPORT NUMBER: TR-76-4

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): 6 UTILITY: RECENT THEORY AND SOME APPLICATIONS.

5. TYPE OF REPORT & PERIOD COVERED: 9 Technical Report, 1976 Interim

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): 10 Marcia D./Deaton

8. CONTRACT OR GRANT NUMBER(s): 15 N00014-76-C-0193

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Decision Making Research
Department of Psychology NI-25
University of Washington, Seattle, WA 98195

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS:
Office of Naval Research Code 452
800 North Quincy Street
Arlington, VA 22217

12. REPORT DATE: 11 September 1976

13. NUMBER OF PAGES: 32 12 41

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Utility Theory, Decision Making, Multi-attribute Utilities, Applied Utility Theory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Recent theoretical advances in multi-attribute utility theory (MAUT) are described, particularly the requirements necessary for its application, and then applications of more or less strict vervions of MAUT are described.

387 783

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Decisions: Recent Theory and Some Applications

"Every human decision does, and should depend on the answers to two questions: What are the odds, and what's at stake?" (Edwards and Guttentag, 1975, p. 416). Thus, the question has progressed from whether expected value or utility theory is useful in decision analysis to what form of utility theory should be used in decision analysis. The purpose of this paper is to look at some of the forms of utility theory, their applications, and the assumptions involved in those applications. This project began as a series of interviews with University of Washington faculty about recommended decision models or strategies in various areas of study. Because many respondents mentioned some variant of utility theory, this paper will focus on it.

The Basic Concepts

Decision making can be divided into four phases: (1) recognition of a decision problem and definition of its nature and dimensions, (2) probability measurement, (3) outcome evaluation, i.e., how good or bad the outcome is, and (4) choice--usually the alternative which has the highest expected value or which returns the most value per unit of cost.

One of the many ways of categorizing decisions is whether they have certain or uncertain outcomes. In the former case, probabilities of occurrence are known to the decision maker, and he can simply insert them into the utility equation. In the latter case, because the probabilities are unknown the decision maker must estimate them; thus, they are subjective probabilities. These subjective probabilities are then used to compute subjective expected utilities.

Ebert and Mitchell (1975) describe expected utility theory as a decision model using beliefs and values. In general, the theory states that an individual's decision can be predicted from his perceptions of the degree to which different alternatives (1) lead to various outcomes (belief/probability) weighted by (2) the evaluation of the outcome (attitudes and values/utilities). The equation for the expected value (utility) rule is:

$$ev = \sum_{i=1}^{N} \Psi_i V_i, \qquad \sum_{i=1}^{N} \Psi_i = 1,$$

where Ψ_i = the probability outcome i will occur,

V_i = the value (utility) of the i^{th} outcome, and

N = the number of outcomes for the decision alternative.

The overall decision rule suggests that the individual will choose the alternative with the highest expected value (utility) in order to maximize his payoffs.

The expected value rule is normative--i.e., it tells you how you ought to behave: If you consistently maximize expected value you will gain more or lose less in the long run than if you don't maximize expected value.

The reasonableness and simplicity of the rule are appealing. "A reasonable theory is one which appeals to you because you agree that you would try to behave in accordance with it, or because if you noticed that you were behaving in a manner that was inconsistent with the theory, you would feel uncomfortable" (Barclay, 1971, p. 6). The expected value rule may not describe exactly the cognitive processes of decision making, "but its acceptance should be based on its predictive success" (Ebert and Mitchell, 1975, p. 61). According to Barclay (1971), expected value theory is a "paramorphic" representation (Hoffman, 1964) of the decision process and need not correspond to subjective experience. "It is not required of models that they bear any semblance of some 'actual' state of affairs, either within the organism or elsewhere, nor would this necessarily lead to a better understanding of nature" (Hoffman, 1964, p. 124).

Ebert and Mitchell (1975) point out two assumptions which are sometimes violated in expected value theory applications. The first is transitivity--i.e., if you prefer A to B and B to C, you will prefer A to C. However, this assumption does not always hold in real world situations. Another questionable assumption of the rule is that beliefs and values are independent of one another.

Theories containing the concept of expected value are prevalent in the areas of learning, personality, attitude formation, decision making, motivation, social power, and leadership (Barclay, 1971; Ebert and Mitchell, 1975; Mitchell and Beach, 1975). Table 1 lists some of these theories.

Insert Table 1 about here

In decision theory, when actuarial probabilities and market values are used to calculate expectations, the term maximization of expected value is used. Values of some outcomes can be stated in objective terms (e.g., dollars), but others require estimates of personal values or utilities (Barclay, 1971). "Utility can be viewed as a private money that allows for 'internal bookkeeping'" (Luce, 1959, pp. 75-76).

When subjective probabilities and subjective values (utilities) are used, the term is maximization of subjective expected utility (SEU). SEU has two variants: expected utility (EU), in which the probabilities are

Table 1

Labels Used for Theoretical Components

Theorist	Determinants of impulse to action
Tolman	Expectancy of goal, demand for goal
Lewin	Potency X Valence
Edwards	Subjective Probability X Utility
Atkinson	Expectancy X (Motive X Incentive)
Rotter	Expectancy, reinforcement value
Vroom	Expectancy X Valence; where valence is Instrumentality X Valence
Peak	Instrumentality X Attitude (affect)
Rosenberg	Instrumentality X Importance
Dulany	Hypothesis of the Distribution of the Reinforcer X Value of the Reinforcer
Fishbein	Probability X Attitude

Note. This table is a modification of one presented by Lawler (1971).

assumed to be 1.00 and are omitted from computations, and weighted expected utility (WEU), in which the probabilities are replaced with an index of importance of each of the various kinds of outcomes under consideration to the decision in general (Mitchell and Beach, 1975).

Using utility theory does not ensure choosing the "right" alternative. The reasoning can be impeccable yet lead to a "wrong" choice in a specific instance. A good illustration is gambling. "One person may play the odds correctly and lose, while another ignores them and wins. Perhaps it may be the case that some individuals have a sixth sense (but this is equivalent to saying that they have more information then we do). Those of us who don't have a sixth sense (or data-collector) can be consoled by the fact that most individuals who do think they have one and who therefore ignore the relevant probabilities eventually suffer the consequences" (Miller and Starr, 1967, p. 81).

Decomposition and MAUT

Because there is serious doubt that man has the ability to process information involving large numbers of dimensions (Shepard, 1964), procedures requiring overall judgments of worth for complex stimuli are unsuitable for many real world problems. Decomposition of the task is recommended, and it can take place at various levels of the decision problem (Humphreys, 1975). Decomposition to level 1 is described as specification of choice alternatives which are usually identified as a set of consequences following from the alternatives. At this level, utilities are assigned to the outcomes, and expected utilities are computed for courses of action.

Thus far much of decision analysis has been concerned with decomposition of decision tasks to level 1--i.e., from an overall decision to a set of judgments about consequences of the various available alternatives. It is possible to decompose to a second level, level 2--i.e., the choice alternatives of level 1 can be broken down into multi-attributed outcomes (Humphreys, 1975). It is here that multi-attribute utility theory (MAUT) becomes a decision aid in decomposing a complex evaluation task into a set of simpler subtasks (von Winterfeldt and Fischer, 1975). Multi-attribute utility models "are designed to obtain the utility of items or alternatives that have more than one valued property and therefore must be evaluated on more than one criterion" (Huber, 1974, p. 1393).

In any of the versions of multi-attribute utility measurement, each outcome to be evaluated is located on each dimension of value. These value location measures are combined using an aggregation rule--frequently a weighted linear combination. The weights describe the relative importance of each dimension of value. "In every application of multi-attribute utilities such numbers are judgmentally obtained" (Edwards and Guttentag, 1975, p. 426).

Although there are a number of multi-attribute utility measurement techniques, those associated with MAUT must satisfy several attribute independence assumptions if they are to be decomposed correctly.

"The degree to which a model allows a decomposition of the evaluation of complex alternatives into independent evaluation aspects such as uncertainties, time discounts, and single attribute utilities distinguishes between the models" (von Winterfeldt and Fischer, 1975, p. 3). That is, choice situations can have either single or multi-attributed outcomes, they can be riskless or risky, and they can be time variant or invariant. The combination of these variables and the results of the independence checks determine the appropriate decomposition model for a choice situation.

In riskless, multi-attributed time invariant choice situations the first check is the Weak Conditional Utility Independence (WCUI) test. If an attribute is WCUI of all others, preferences for values in that attribute are independent of constant values in the other attributes. That is, do other dimensions interfere with your ability to maintain the monotonic preference order of a dimension? If not, that dimension is 1-WCUI of the others. If 1-WCUI fails, i.e., if no attribute is WCUI of the others, the choice is left to the decision maker's intuition. If 1-WCUI holds, the next test is for n-WCUI, i.e., if 1-WCUI holds for all n attributes, if this test fails, then outcomes which are equated on all attributes except one are compared, and choices are prescribed on the basis of the monotonicity of the utility function over this attribute.

If n-WCUI is satisfied, joint independencies are tested. "A set of attributes is said to be jointly independent of the rest if the preference order of alternatives which vary only in these attributes remains invariant for any fixed levels of the remaining attributes" (von Winterfeldt and Fischer, 1975, p. 40).

If no joint independence condition is satisfied, a total decomposition is applied which allows the construction of independent utility functions in the single attributes. With this model trade-offs between some dimensions independent of the values of others can be considered and alternatives which are dominated in all single dimension utilities can be excluded.

If the assumption of joint independencies is satisfied, single attribute utility functions can be constructed, and the sum of these utility functions represents the worth of each alternative.

in risky, time invariant, multi-attributed choice situations, there is another set of assumptions. One, the sure thing principle, says that preferences among gambles should not depend on the values of outcomes which are constant in a subset of events. This assumption is similar to the joint independence assumption and can be checked by constructing some critical examples rather than testing all subsets of events and all combinations of outcomes.

A second assumption in an expected utility representation is that no outcome should be infinitely desirable or undesirable. "Thought experiments" (von Winterfeldt and Fischer, 1975, p. 27) are usually sufficient for checking this assumption.

If either of these assumptions is violated, MAUT and decision theory in general, can be of little help to the decision maker. Von Winterfeldt and Fischer (1975) offer Coomb's protfolio theory and the minimax models as possible, but unsatisfactory, alternatives to MAUT in this situation.

However, if these assumptions are satisfied, the next test is Strong Conditional Utility Independence (SCUI)--the probabilistic equivalent of WCUI and joint independence. "SCUI says that preferences among uncertain multi-attributed alternatives in which a subset of attributes has constant values across all outcomes should not depend upon the particular level at which these constant values are held fixed" (von Winterfeldt and Fischer, 1975, p. 28).

If SCUI does not hold, it is necessary to go through the tests of assumptions for riskless situations to determine the appropriate model of decomposition. There are means available for transforming riskless utility functions into an expected utility representation (see Raiffa, 1969).

If SCUI is satisfied, the next, most stringent test of independence is of marginal equivalence or marginality. This test requires that risky multi-attributed alternatives be judged only on the basis of expected utilities. That is, if two or more events have equal expected utilities, shuffling the component outcomes should not influence their desirability. For example, if marginality holds, the following gambles should be equally attractive:

$\underline{X}$ =
- $p = .5$ → \$4000 + a 1973 Porsche
- $(1-p) = .5$ → \$0 + a 1961 VW

$\underline{Y}$
- $p = .5$ → \$0 + a 1973 Porsche
- $(1-p) = .5$ → \$4000 + a 1961 VW

In this case, both $\underline{X}$ and $\underline{Y}$ have the same expected value, but because most people would prefer the right bet to the left, the two gambles are not marginally equivalent.

If marginality is not satisfied but SCUI is, the choice situation can be represented by a multiplicative expected utility model. Alternatively, if both marginality and SCUI assumptions are satisfied, an additive expected utility model is justified.

Edwards (1975) notes that there is a wide gap between the theoreticians and users of MAUT. Users are more concerned with applications than with independence assumptions and underlying measurement theory. Although decision analysis is relatively insensitive to the error of applying a theoretically inappropriate model, users should be aware of the implications of the various models and understand the relationship among the models.

For cases in which independence assumptions are violated, Humphreys (1975) describes three alternatives available to the decision maker or analyst: (1) a partial decomposition model can be applied (as recommended by von Winterfeldt and Fischer, 1975), (2) attribute dimensions can be adjusted to allow total decomposition, and (3) the violations can be ignored and a total decomposition model applied. The last alternative, the forced decomposition solution, is the most common approach.

Edwards and Guttentag (1975) describe one method of using MAUT. In identifying the relevant dimensions of value, it is important not to be too expansive. The list need not be exhaustive--in fact, it should include a few important, general values and omit the less important goals. The dimensions should then be ranked and rated on the basis of importance. These importance weights (ratings) are transformed to a 100 point scale. Each entity being evaluated is then measured on each dimension--i.e., a sort of performance measure of each alternative is obtained for each dimension. These measures are converted to a 100 point scale (usually by a linear transformation) on which 0 is the minimum plausible performance and 100 is the maximum.

The next step is to calculate utilities for the entities by summing the products of the normalized importance weights and the rescaled position of the entity on each dimension--i.e., compute a weighted average. The decision rule is to maximize utility. In cases where there are budget constraints, the benefit-to-cost ratio should be maximized.

A matrix can then be generated with the entities being evaluated as rows and the value dimensions as columns. The performance measures fill the cells. As more data are accumulated, initial guesses can be updated and revised. Thus MAUT offers a process for evaluating programs as well as planning them.

Decomposition and Decision Trees

The decomposition of a decision problem can be graphically represented by decision trees and hierarchies. The decision tree, or decision flow diagram, presents "the anatomy or the qualitative structure of the problem as a chronological arrangement of those choices that are controlled by [the] decision maker, and those choices that are determined by Chance" (Raiffa, 1968, p. 10) as well as the consequences of those choices. That is, a decision tree is a stylized roadmap of a problem.

To continue the metaphor, the tolls you must pay are marked on the map, and the penalties or prizes are designated at the terminal ends (tips of the branches). Probabilities are computed or estimated (depending on the certainty of the outcome) for each of the branches. The averaging out process involves summing the products of the probabilities and the payoffs at each chance juncture and the folding back procedure involves selecting the option which maximizes expected value at each decision juncture. Circles are used to signify chance junctures or event forks and squares represent decision nodes or act forks (Brown, Kahn, and Peterson, 1974).

Both decision trees and hierarchies can be used to represent single and multi-attribute problems. A hierarchy, too, shows the anatomy of a decision problem, but not necessarily as a chronological arrangement of choices. It may simply reflect the breaking down of a goal or outcome into its relevant components. Utilities and probabilities can be computed or estimated for the components as they are for the decision tree.

A Potpourri of Examples

Some examples of applications of single and multi-attribute utility theory may help clarify them and give the reader a feeling for their range of usefulness.

<u>Career preferences</u>. Holmstrom and Beach (1973) used an SEU approach to study career preferences of senior psychology majors intending to go to graduate school. Eight psychological occupations were decomposed to 18 kinds of payoffs or outcomes. Subjects ranked the occupations in order of their preferences and then arranged the ranked occupations on a 100 point scale, placing the least preferred occupation at 0 and the most preferred at 100. The remaining occupations were distributed along the scale to indicate their relative preferability.

In order to determine utilities, subjects then ranked the 18 kinds of payoffs in terms of the relative importance that they be satisfied by a future career. Then they arranged the payoffs on a 100 point scale of importance with the least important payoff at 0, the most important at 100, and the others along the scale to describe their relative importance. To determine subjective probabilities, subjects estimated the probability that each of the 18 kinds of payoffs would be satisfied by each of the eight occupations. SEU's were computed for each of the eight occupations for each subject on the basis of his utilities and subjective probabilities. That is,

$$SEU = \sum_{i=1}^{18} [P_iU_i + (1-P_i)(-U_i)], \text{ where}$$

P_i = the stated probability of success of payoff i,

U_i = the utility of payoff i,

$1-P_i$ = the probability of failure of payoff i, and

$-U_i$ = the difficulty or disutility of payoff i,

For each subject a correlation was computed between the relative preferences for the eight occupations and the eight computed SEU's associated with the occupations. The results show that relative occupational preference can be accounted for by the relative magnitudes of the SEU's for the occupations.

"For convenience, the 18 payoffs were assumed to be independent. Although it is unlikely that this assumption is correct, results from previous research (e.g., Hoffman, 1960) suggest that the violation of independence need not be too misleading and, as it turned out, the results of this study indicate that no great damage was done by the assumption" (Holmstrom and Beach, 1973, p. 203).

Education. Educational assessment has repeatedly been hindered because there has been no technical procedure for measuring values and growth. Page (1972) proposed a measure of general educational advancement called the "bentee"--a "benefit T-score for education." The bentee is a normalized, equal-interval scale adjusted to a norm of some comparison group (e.g., high school seniors). It has a mean of 50 and a standard deviation of 10. The bentee can be subdivided into seven traits. Each trait becomes a node in the developing tree, and each node has its own branches which lead to other nodes, as shown in Figure 1.

Insert Figure 1 about here

Page (1974) notes that the nodes and terms are matters for discussion by expert panels. According to Page, "the bentee provides for a reasonable movement from democratic principles to technical expertise" (p. 576). That is, at the top of the hierarchy, the judges may be citizens, students, parents, board members, or elected officials. Judges at the lower levels, however, are likely to be psychological or subject-matter experts.

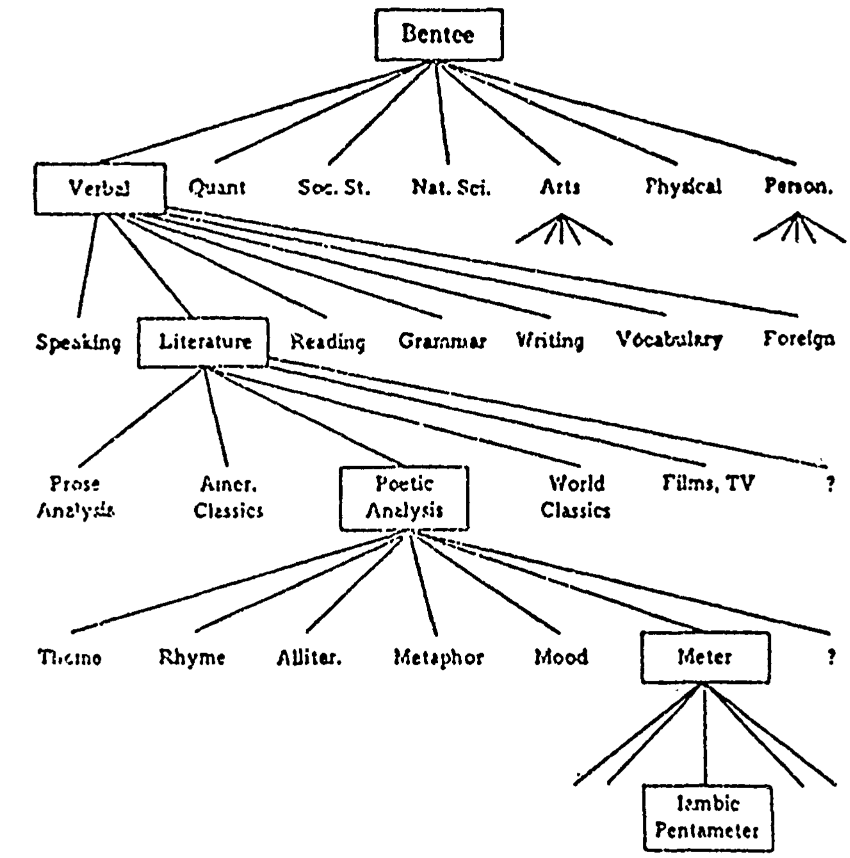

Figure 1. The bentee hierarchy.

Once the hierarchy has been developed, weightings for the nodes are established by one of two means. Using the token system, each judge "spends" 100 tokens or poker chips to indicate his judgment of each subtrait's importance to the node above it. The number of tokens spent on each branch reflects its proportion of importance within the node-family. The branches of each node within each level of the hierarchy must sum to 1.00. The resulting scale of weightings is considered a ratio scale; no chips would indicate no importance, and 40 chips would indicate twice as much importance as 20 chips. The second method for establishing weightings is the correlational strategy. Each judge is presented with computer simulated profiles of sample "students." Each profile consists of the student's T-scores on the traits at one level of the tree. The judge is asked to assign an overall value to the profile to indicate his assessment of how well-educated the student is. The profiles are randomly generated, and there is no deliberate co-variance among the traits. Because the criterion correlations are "direct measures of the latent evaluation of the traits, they are equivalent of beta weights, and may serve as coefficients of the measures in any student or group profile for calculating the bentee for that student or group" (Page, 1973, p. 23).

The weights are averaged across judges in order to establish the values of society or of a profession. To appraise a student's education,

$$\text{Bentee}_j = T\left(\sum_{i=1}^{n} v_i m_{ij}\right),$$

where v_i is the societal value of the i^{th} trait, m_{ij} is the measure (in standard scores) of the i^{th} trait for the j^{th} student, and T is the transformation of the sum of weighted measures to the T distribution. This technique is applicable at each node in the value tree. The values of the branches sum to 1.00 within each node family, and the value of each node is the product of all the lineal values above it.

Medical diagnosis. Yondorf (1972) proposed the use of decision trees in the analysis of which females to test, retest, and treat for gonorrhea. She estimated probabilities of disease occurrence on the basis of her knowledge of the literature on gonorrhea and then assigned subjective utility values to reflect her judgment of the cost of gonorrhea. Yondorf constructed three decision trees--one for a woman who is coming in for an examination and who has all the symptoms of the disease, a second for a woman who has no symptoms but whose sexual partner(s) has gonorrhea, and a third for a woman who is coming in because of a different complaint but who lives in a high incidence of gonorrhea area as defined by the Department of Health, Education, and Welfare. Figure 2 shows the tree for the first case.

Insert Figure 2 about here

Lusted (1971) proposed a variation of MAUT to help the physician choose a course of action taking into account his personal judgments about probabilities, costs, and his preferences for consequences of diagnoses and treatments. Lusted uses signal detection theory and receiver-operating characteristic (ROC) curves to determine the relative weight a physician attaches to the values of correct diagnoses compared with costs of errors. An ROC curve is plotted on two independent quantities--the percentage of true positive (TP) and the percentage of false positive (FP) diagnoses made by a physician in a series of "proved cases" where the correct diagnosis is known to the experimenter. A likelihood ratio (the ratio of two conditional

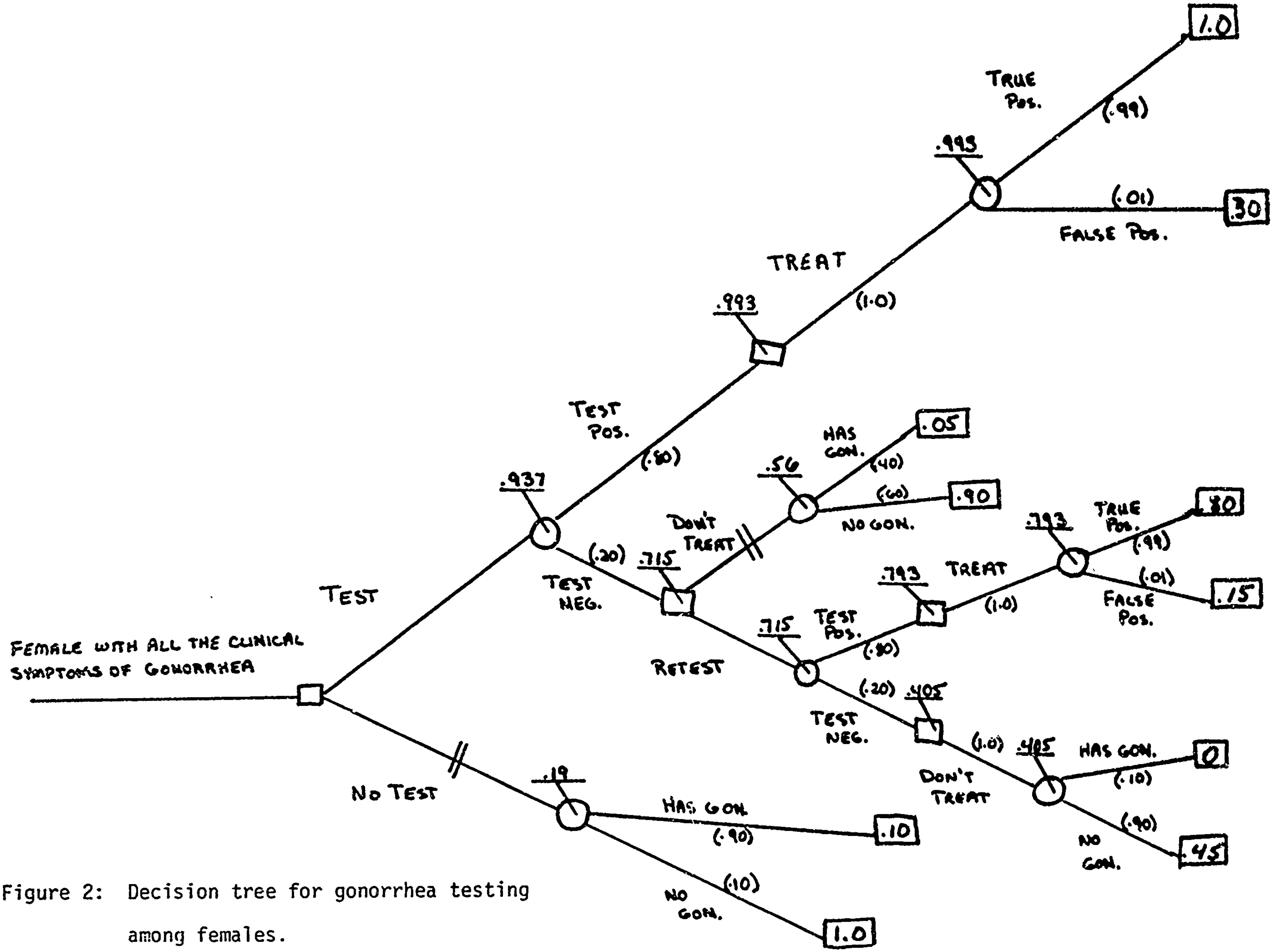

Figure 2: Decision tree for gonorrhea testing among females.

probabilities) can be determined from the ROC curve. In turn, an expected value of a diagnosis can be determined from the likelihood ratio and the values and costs associated with an outcome. The equation for the expected value is:

$$L_c = \frac{p(TN)}{p(TP)} \times \frac{V_{TN} + C_{FP}}{V_{TP} + C_{FN}}, \text{ where}$$

L_c = cut off or critical value of criteria on the basis of which the physician makes his decision,

$p(TN)$ = probability of a true negative diagnosis,

$p(TP)$ = probability of a true positive diagnosis,

V_{TN} = value (for the physician) of a true negative diagnosis,

V_{TP} = value (for the physician) of a true positive diagnosis,

C_{FP} = cost (for the physician) of a false positive diagnosis, and

C_{FN} = cost (for the physician) of a false negative diagnosis.

The slope of an ROC curve at any point is equal to the likelihood ratio criterion that generates that point.

The physician then answers a series of questions about his values. However, instead of requiring an explicit expression of values, Lusted's method asks the physician for his attitude about the relative value of true positive vs. true negative diagnoses ($\frac{V_{TP}}{V_{TN}}$). The ratio can be 1, < 1, or > 1. The physician answers questions about the relative cost of diagnostic errors compared with the value of correct diagnoses.

In an example using his own responses in an experiment, Lusted concluded that for him, for the diagnosis of active tuberculosis, $V_{TP}=V_{TN}$ (i.e., $\frac{V_{TP}}{V_{TN}} = 1$) and $C_{TN} >> 1$ and C_{FP} 1. His likelihood ratio ($\frac{p(TN)}{p(TP)}$) equals 2000, and the slope of the ROC curve is 5 at the operating point of 80 percent true positive and 4 percent false positive diagnoses. Thus, if $\frac{V_{TP}}{V_{TN}} = 1$, $\frac{p(TN)}{p(TP)} = 2000$, $L_c = 5$, $C_{FN} >> 1$, and $C_{FP} >> 1$, using the previous equation,

$$5 = 2000 \cdot \frac{C_{FP}}{C_{FN}}\text{, or } C_{FN} = 400(C_{FP}).$$

That is, the cost of a false negative diagnosis must be 400 times the cost of a false positive diagnosis to warrant a positive diagnosis of tuberculosis. That means that "the consequences of ignoring tuberculosis when it is present must be at least 400 times as serious as the consequences of further testing or treatment of tuberculosis when the patient does not in fact have tuberculosis" (p. 419).

Water quality assessment. O'Connor (1973) applied multi-attribute utility scaling procedures to the assessment of water quality for the public water supply and for fish and wildlife. He used a procedure proposed by Edwards (1971) involving ratio scale judgments of interdimensional importance weights, single dimensional utility functions, and the use of an additive combination rule. However, he pointed out that in the complex task of assessing water quality, the additive model was unable to handle "the configurality involved in water quality chemistry. Side conditions of a conjunctive nature were created to capture the complicated environmental interactions among parameters as well as the toxicities of parameters in extreme ranges" (p. 35).

O'Connor emphasizes that the criterion with respect to which value judgments are to be made must be very clear and also agreed upon by the judges. Lack of consensus on the importance weights was not a crucial factor in determining the values assigned to samples of water by individual indices. As predicted by the literature on the robustness of the additive model, the indices were insensitive to variations of value functions and parameter weights. According to O'Connor, "more effort should be placed on methods of obtaining group consensus on the structuring of problems, decision criteria, and dimensions of value rather than obtaining consensus on estimates of parameters and/or utility functions" (p. 34).

The conclusions drawn from this study are (1) that MAU assessment procedures are applicable to real world problems, but (2) "multi-stage models involving sequential application of various types of decision rules (compensatory and noncompensatory) are likely to be necessary in any valid quantification of real world problems" (p. 36).

Urban transportation. One of the recently proposed approaches for aiding urban transportation planners and engineers evaluate a set of plans and recommend a course of action is a worth assessment procedure for plan alternative comparison. The object of this technique, originally proposed by Miller (1970), is to develop "a numerical index to measure the complete worth of a set of complex alternatives, given specific social values for the utility of achieving levels of performance on evaluation attributes" (Parsons, 1973, p. 2). Performance on each attribute is converted to an abstract unit of measure, the importance of the attributes is weighted, and the products of the performance worths and weights are summed to yield a total worth score. Because this method enables the analyst to derive a quantitative measure of the value of various alternatives from individuals' specific value-sets, it is particularly well suited to public planning decisions.

The procedure developed by the Urban Systems Research Center at the University of Washington is based on the Miller approach. In the first step of implementation, tentative general goals are developed in all relevant goal categories of a plan. These may be in the form of slogans such as "maximize the level of service provided by a given system," or "minimize the environmental impact of a certain action." From these goals a set of upper level criteria, or factors by which an alternative will be evaluated, is formulated. The upper level criteria are subdivided to form a hierarchy of worth, also called an evaluation tree. In this tree, each lower level element contributes to the worth of the element above it. This subdividing continues until a list of attributes is established, for which performance measures can be specified.

Performance measures and units of measure for each attribute are selected. In situations where performance for certain attributes cannot be quantified easily, decision makers are required to subjectively evaluate each plan and convert performance to worth units directly. This direct worth estimate is made, for example, in the evaluation of some aspects of environmental impact such as aesthetics or neighborhood disruption. Then projected levels of performance are developed for each performance level of each plan.

Worth curves are developed so that levels of performance for various attributes can be scaled over a common range of units. Scales of 0 to 10 or 0.0 to 1.0 are typically used, with the low point representing very poor performance and the high point representing an absolute optimal value. It is necessary to generate a set of worth curves for each decision maker because the worth of any given level of performance is dependent upon the value-set of the individual making the assessment. Each decision maker is asked to specify an Ideal Standard (the level of performance beyond which further improvement will bring only a minimal increase in system worth) and an Acceptable Limit (the lowest or highest level of performance that will be tolerated. The specification of these two points on the curve determines the exact relation between performance and worth.

Next a matrix of unweighted worth scores is developed for each plan. The matrix, consisting of scores on each performance measure (PM) for each individual (I), will have dimensions PM x I. Another matrix of the same dimensions is developed using the weights for each attribute as specified by each individual. There are several methods of determining weights. Ranking, scaling, and paired comparisons are possibilities, but, in experiments conducted at the University of Washington, the hierarchical comparisons approach was found to be the most workable process. This method is based upon the evaluation tree developed in an earlier stage of the assessment procedure. An important assumption is that at each branch of the tree all the factors contributing to an element of a higher level have been identified. Decision makers are asked to rate the relative importance of the contribution of each lower level element to the element above by dividing a constant sum among them. Weights are calculated by starting at the top or the bottom value of the tree and forming a product of the values that appear at each branch as one progresses up or down the tree.

Finally, the products of the individual performance measures and weights are summed to produce a total worth score for each alternative which represents the judgments made by each individual. "A means for analyzing the total worth scores to determine the preferred alternative must then be decided upon" (Parsons, 1973, p. 30).

The comparative worth approach is viewed as a cyclical technique. In the experiment at the University of Washington, decision makers received feedback about their own and their group members' judgments and were given the opportunity to discuss areas of conflict and then change their judgments.

Birth planning. Beach, Townes, Campbell and Keating (1976) have developed a hierarchy based on multi-attribute utility measurement as an aid for birth planning decisions. Obviously in this choice situation there are two alternatives--either the couple has a(another) child or it does not. These alternatives involve outcomes involving the values centered on self and spouse, children, and significant others. Each of these outcomes is multi-attributed. As Figure 3 shows the attributes of values centered on self and spouse, for example, are personal identity, parenthood, and well-being of the family. In this hierarchy, each attribute is accompanied by a non-exhaustive list of exemplars which suggest to the decision maker what items could be considered at the lowest level of the hierarchy.

Insert Figure 3 about here

Instead of addressing the specific independence assumptions discussed by von Winterfeldt and Fischer (1975), Beach et al. impose 3 conditions on their hierarchy. First, at any level of the hierarchy, each item is a constituent of only one category, and "sub-categories are in turn constituents of one category on the next level of the hierarchy" (Campbell, Townes and Beach, p. 6). The second condition is that constituents of a class must be exhaustive and independent of one another. Third, evaluations of the constituents may range from .00 to 1.00, but they must sum to 1.00 over all constituents of the class.

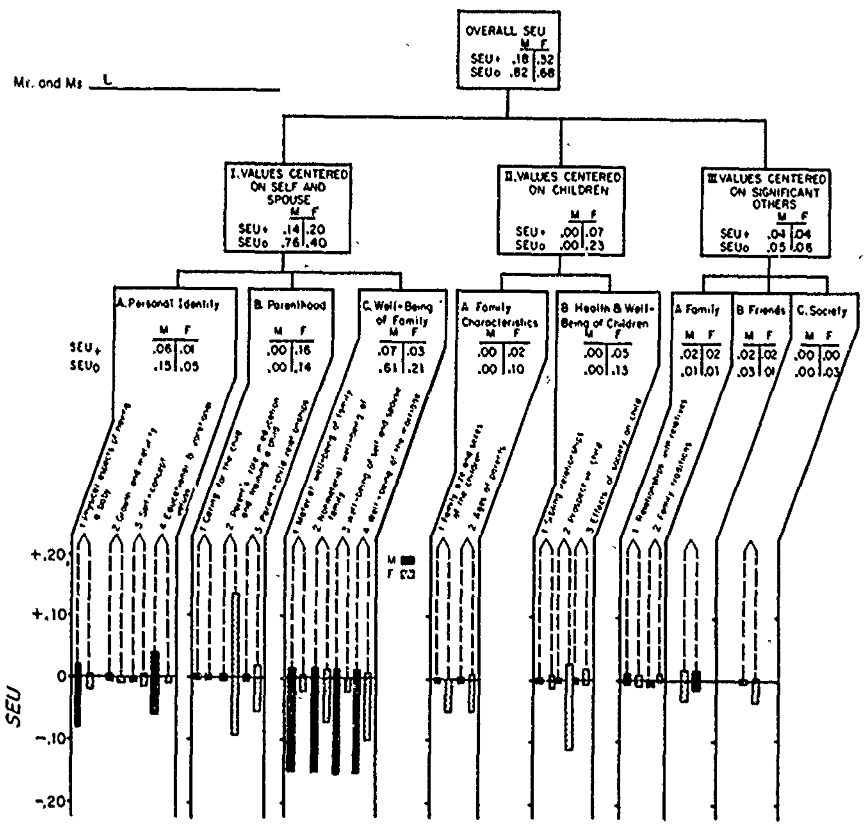

Figure 3. An example of a couple who wish to remain childless.

Beach et al. recognize the impossibility of fully satisfying these conditions. "But, if some reasonably close approximation can be achieved, the scheme may prove sufficiently useful in practice to justify indulgence of its imperfections" (Beach et al., 1976, p. 103).

With these three conditions satisfied, the decision maker can decompose the hierarchy into a series of separate decisions about the importance and probability of occurrence of the constituents of any class. Starting at the top of the hierarchy, the decision maker rates the importance of each of the classes in accordance with the third condition discussed previously. He then multiplies down each path of the hierarchy to the bottom. When this process is completed, he assigns a + or - sign to each of the resultants, indicating whether he would view the class of outcomes as a gain or a loss if a child were born.

For each bottom level class the decision maker estimates his subjective probability, p, that the outcomes that define that class would occur if he were to elect to have a child. Note that this also determines the probability, 1 - p, for not having a child. The SEU's for having and not having a child are computed by multiplying the subjective probabilities and utilities and summing across all of the bottom level classes. The decision maker should choose the alternative with the larger SEU, and the greater the difference between the two SEU's, the more clear-cut the decision.

In conventional use of decision trees, probabilities are located at the nodes, and utilities appear at the ends of the branches (Raiffa, 1968). This practice reflects the nature of many tasks--i.e., uncertainty of future events is emphasized, and decomposition is aimed at simplifying probability assessment. In the birth planning task, the emphasis is reversed because utility considerations are harder to sort out and keep track of than are probabilities. Thus the conceptual clarity afforded by the use of trees can be put to work for utility analysis.

There are two possible uses for this application of multi-attribute utility measurement to borth planning decisions. First, it may have predictive value, and second, used in a counseling situation, it may help couples to clarify their pre-decisional thinking and "base their decisions on a calculus of conscious choice" (Campbell et al, 1976, p, 22).

Urban planning. Fitzpatrick (1974) has reviewed several methods of proposal testing, some of which make use of MAU measurement procedures. For example, Hill (1968) presents the goals-achievement matrix as a method of proposal testing. "The key to plan evaluation by means of goals-achievement analysis is the weighting of objectives, activities, locations, groups, or sectors in urban areas" (p. 27). Relative weights are established for an explicit set of goals, and objectives are defined in operational terms. For each objective, the consequences of each alternative plan are determined. Each alternative is then measured in terms of its achievement of each goal. The product of the analysis is presented in a summary table for each alternative, and the summation of the evaluation of several alternatives can be presented in a weighted Goals-Achievement index.

This approach is based on benefit-cost analysis but emphasizes defining a specific set of goals and objectives against which to judge each alternative. Benefit-cost analysis has traditionally concentrated on one general goal--economic efficiency.

Recognizing the existence of limitations on the usefulness of these techniques and the quality of information, Hill advocates using (1) a variety of methods of determining objectives to check for complementary findings, and (2) sensitivity tests to check the degree to which changes in variables may affect the results. He suggests the use of probability theory to deal with the uncertainty in predicting and evaluating the consequences of alternatives, and he recommends that the Goals-Achievement Matrix now be used only for the evaluation of plans in a single sector because the analysis cannot yet account for interactions between objectives.

A very similar approach is Schimpeler and Grecco's (1968) Community Structure and Values model designed for transportation planning. Community goals and objectives are listed, and utility values are established for each objective. These utility values may vary for different socio-economic groups in the community. Each proposal is judged by each group on the basis of how well it fulfills each objective. Each group's score for an alternative is the sum of the products of the utility value for each objective and an effectiveness score ranging from 1.0 (indicating complete fulfillment of an objective) to 0.0 (for virtual impossibility of fulfillment). The utility value for each alternative is computed by summing the groups' utility scores for that alternative.

Fitzpatrick included in her review Vesper and Sayeki's (1972) Successive Weighting method which involves a hierarchy with objectives at the top level, policy areas in the middle, and alternative actions at the bottom. First, 10 points are divided among the objectives to indicate their relative importance to goal achievement. A second set of 10 points is divided among policies on the basis of their relative strength in fulfilling each objective. A final set of 10 points is divided among alternative actions within a policy to indicate their relative effectiveness in implementing the policy. To evaluate each action, assigned weights are multiplied from the bottom up, i.e., the weight for an alternative is multiplied by the weight for the policy, and this product is multiplied by the weight for the objective. For a set of actions, the utility scores for each action are added together. The highest score, then, indicates the best single alternative or set of actions.

There are several assumptions that are critical to this method: (1) lists of relevant policies and objectives should be comprehensive, but the list of actions need not be, (2) objectives and policy areas should be defined as independently as possible and should not semantically duplicate or overlap each other, (3) measurement scales should be consistent throughout the analysis, and (4) all actions within a policy area distribute their influence on achievement of objectives in the same proportions.

The technique proposed by Langley (1973) for evaluating the effectiveness of alternatives in structure planning involves summing the products of (1) the performance of an alternative in achieving a specific objective, and (2) the relative importance of a specific objective. There are many methods of determining the weights and performance scores (see Langley, 1973, 1974).

Dearborn (1970) has developed a decision technique for multi-dimensional decisions between discrete or mutually exclusive courses of action in the daily decision environment. He rejects classical decision theory on the grounds that it is unrealistic and its applicability is very narrow. The decision environment is frequently uncertain, access to information is limited, and all possible alternatives are rarely identifiable; the decision maker's orientation in such an environment is necessarily subjective. According to Dearborn, under these conditions the decision maker cannot maximize in trying to satisfy his objectives, so he often "satisfices" (Simon, 1957)--i.e., selects an alternative that is "good enough." The decision process as described by Dearborn appears as Figure 4.

Insert Figure 4 about here

In the first step, goals will be qualitative (achievable or not achievable) or quantitative (obtainable in various degrees). The goals and subgoals or evaluation factors can be arranged in a hierarchy. The qualitative goals, translated into operational terms, can be used to define a limited number of feasible alternatives. Consideration of all possible alternatives is not necessary when satisficing. Depending on the number of alternatives the search yields, it may be necessary to return to step 1 and broaden the goals or set higher standards. The consequences or outcomes of each alternative are then determined and measured. An ideal goal ranking based on the degree of goal achievement is used as a standard of comparison for each alternative.

Frequently, outcomes do not mutually support one course of action; each alternative may be favored for one of several goals. When goals conflict, in order to make a choice, the decision maker must balance one goal against another, and goal weights are used to place goals in their proper (according to the decision maker) relationship to each other. The weights of the goal-achievement ratings can be added to obtain a rank order. If the resultant ranking does not significantly favor one alternative, the outcomes must be reassessed and the consequences re-evaluated.

Dearborn applies this decision process to the Sentinel anti-ballistic missile urban site selection procedure. In this case, the site selection

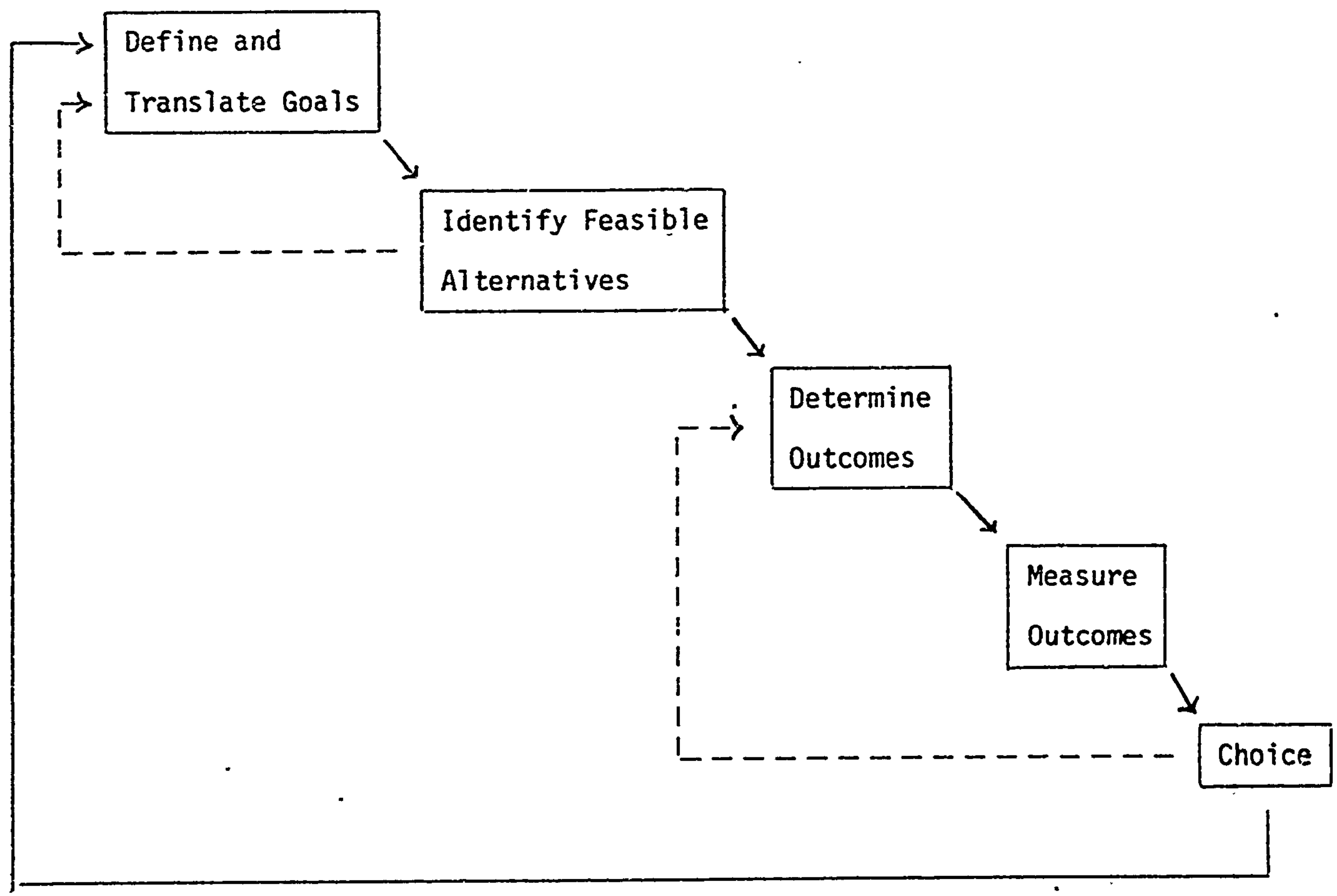

Figure 4: Dearborn's Decision Process Model.

decision process began with feasible alternatives identified by the Army. One of the qualitative goals used to determine the feasibility of alternatives was the defensibility of the location. This goal can be operationalized as the site's having x pounds per square inch of soil bearing capacity and y degrees of radar scanning potnetial. If a site lacked these capacities, it was not considered a feasible alternative and was eliminated from the analysis.

A hierarchy of the quantitative goals was developed (Figure 5). One of the quantitative goals, minimizing installation and operation costs, was operationalized as land and construction cost, among other things.

Insert Figure 5 about here

The next steps were to evaluate each alternative's degree of goal achievement and assign importance weights to rank goals. There are a number of methods of determining these weights and developing rankings. If the measurement scales cannot be defined in similar units, then each goal should be quantified in terms most convenient for its assessment and then converted to a common scale through a transformation function to yield a final comparative value. When alternatives are closely ranked, a determination of weight sensitivity is important to eliminate the arbitrariness of weights.

Once the goal achievement values and importance weights were determined for the Sentinel project, a site value was calculated for Tier I by summing the products of the installation value x installation weight and community impact value x community impact weight, and then multiplying this sum by the tactical value of the site. This computation yielded an effectiveness-to-

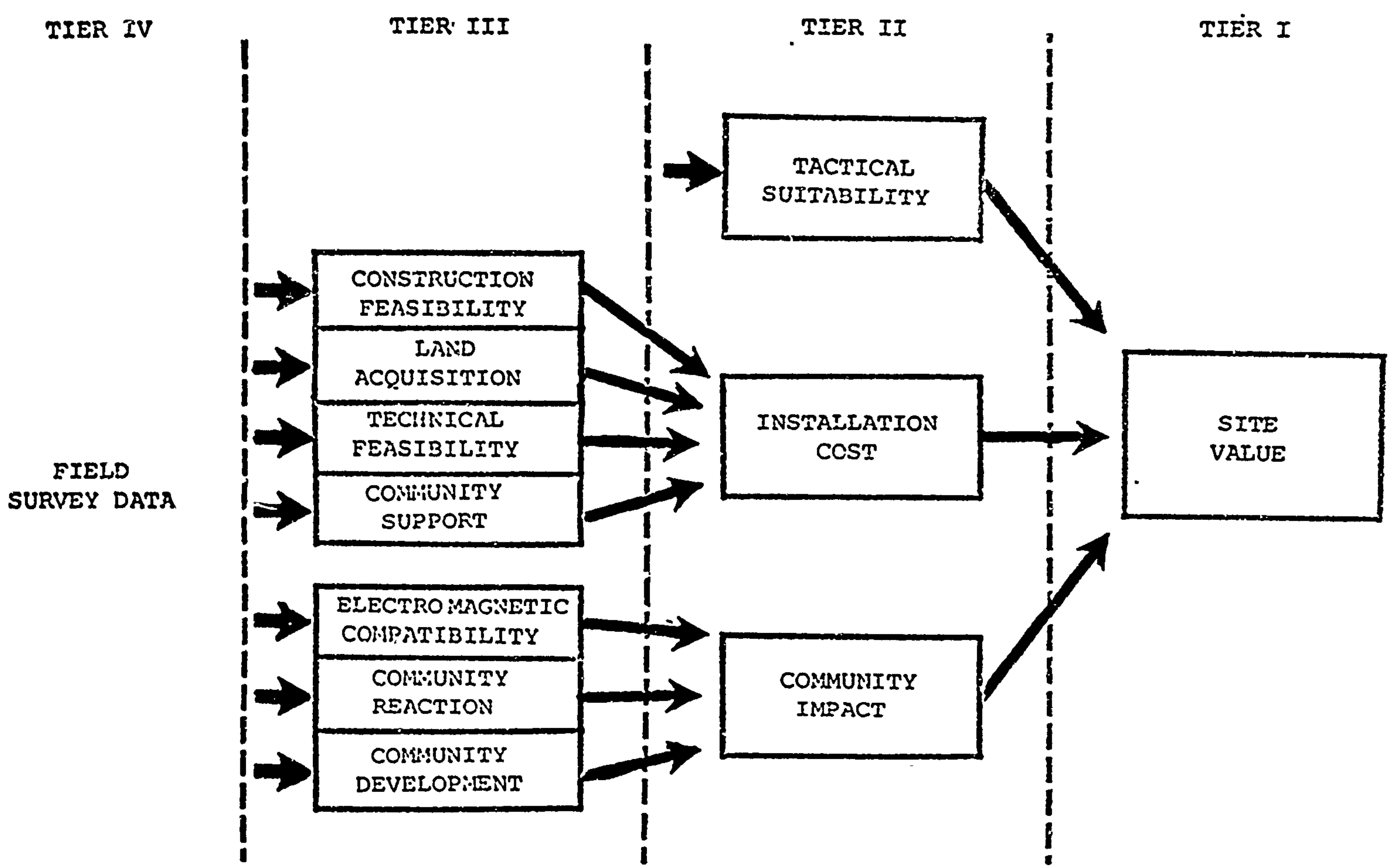

Figure 5: Sentinel goal hierarchy.

cost relationship among the major factors; as a site's tactical value increases or decreases, its overall value should change proportionately. This logic does not hold for the other two factors. For example, a site's value may be high if it protects urban areas, but this value may be inversely proportional to the installation or impact value. For the other tiers, values of factors were determined by summing the products of the importance weights and the goal achievement ratings. Because this decision was to be made by a group, compromises were made until, in the opinion of all participants, a satisfactory weight relationship had been determined.

References

Barclay, S. Expected value in psychological models. Unpublished manuscript, March, 1971.

Beach, L. R., Townes, B. D., Campbell, F. L., & Keating, G. W. Developing and testing a decision aid for birth planning decisions. Organizational Behavior and Human Performance, 1976, 15, 99-116.

Brown, R. V., Kahr, A. S., & Peterson, C. Decision analysis: An overview. New York: Holt, Rinehart and Winston, 1974.

Campbell, F. L., Townes, B. D., & Beach, L. R. Counseling for child-bearing: Toward a calculus of conscious choice. Unpublished manuscript, 1976.

Dearborn, K. W. A weighted-value approach to decision-making. Unpublished master's thesis. University of Washington, 1970.

Ebert, R. J., & Mitchell, T. R. Organizational decision processes. New York: Crane, Russak, 1975.

Edwards, W. Social utilities in decision and risk analysis: New tools for management. Proceedings of the Sixth Triannual Symposium, Hoboken, June, 1971.

Edwards, W. Research on the technology of inference and decision (Report Number 75-10). Los Angeles: Social Science Research Institute, August, 1975.

Edwards, W., & Guttentag, M. Experiments and evaluations: A reexamination. In C. A. Bennett and A. A. Lumsdaine, Evaluation and experiment. New York: Academic Press, 1975, pp. 409-463.

Fitzpatrick, L. Evaluating alternative plans, current methods and a proposal. Unpublished master's thesis, University of Washington, 1974.

Hill, M. A goals-achievement matrix for evaluating alternative plans. Journal of the American Institute of Planners, 1968, 34, 19-28.

Hoffman, P. J. The paramorphic representation of clinical judgment. Psychological Bulletin, 1960, 47, 116-131.

Holmstrom, V. L., & Beach, L. R. Subjective expected utility and career preferences. Organizational Behavior and Human Performance, 1973, 10, 201-207.

Huber, G. P. Multi-attribute utility models: A review of field and field-like studies. Management Science, June, 1974, 20, 1393-1401.

Humphreys, R. Applications of multi-attribute utility theory. Paper presented at the Fifth Research Conference on Subjective Probability, Utility and Decision Making. Darmstadt, September, 1975.

Langley, P. Techniques for evaluating the effectiveness of alternatives in structure planning. D.O.E. Planning Techniques Paper No. 8, April, 1973.

Langley, P. E. Methods of defining and assessing the importance of structure plan objectives (DPRS 3 Planning Techniques Paper 2/74). London: Department of the Environment, December, 1974.

Luce, R. D. Individual choice behavior. New York: John Wiley, 1959.

Lusted, L. Decision-making studies in patient management. New England Journal of Medicine, 1971, 284 (8), 416-424.

Miller, D. W., & Starr, M. K. The structure of human decisions. Englewood Cliffs, N.J.: Prentice-Hall, 1967.

Miller, J. R., III. Professional decision-making. New York: Praeger, 1970.

Mitchell, T. R., & Beach, L. R. Expectancy theory, decision theory and occupational preference and choice. (Technical Report 75-75). Seattle: University of Washington, December, 1975.

O'Connor, M. F. The application of multi-attribute scaling procedures to the development indices of water quality. (Report number 7339). Chicago: University of Chicago Center for Mathematical Studies in Business and Economics, 1973.

Page, E. B. Seeking a measure of general educational advancement: The bentee. Journal of Educational Measurement, 1972, 9, 33-43.

Page, E. B., & Breen, T. F., III. Educational values for measurement technology: Some theory and data. In W. E. Coffman (Ed.), Frontiers of educational measurement and information systems. Boston: Houghton-Mifflin, 1973.

Page, E. B. "Top-down" trees of educational values. Educational and Psychological Measurement, 1974, 34, 573-584.

Parsons, J. D. An analytic methodology for the evaluation of multi-attributed transportation planning alternatives under multi-goaled objectives. (Research Report Number 73-5). Seattle: University of Washington Urban Transportation Program, November, 1973.

Raiffa, H. Preferences for multi-attributed alternatives. RAND Memorandum 5868. Boston, 1969.

Raiffa, H. Decision analysis. Reading, Mass.: Addison-Wesley, 1968.

Schimpeler, C. C., & Grecco, W. L. The community: Systems evaluation: An approach based on community structure. Highway Research Record, 1968, 238, 123-152.

Shepard, R. N. On subjectively optimum selections among multi-attribute alternatives. In M. W. Shelley and G. L. Bryan. Human judgments and optimality. New York: Wiley, 1964, pp. 257-281.

Simon, H. A. Models of man. New York: Wiley, 1957.

Vesper K., & Sayeki, Y. A quantitative approach for policy analysis. Unpublished manuscript. Seattle: University of Washington, 1972.

von Winterfeldt, D., & Fischer, G. W. Multi-attribute utility theory: Models and assessment procedures. In D. Wendt and C. Vlek (Eds.), Utility, probability and human decision making. Dortrecht: Reidel, 1975.

Yondorf, B. A model for gonorrhea testing among females. Unpublished manuscript, University of Washington, 1972.

FOOTNOTE

[1]This research was partially supported by Office of Naval Research Contract N00014-76-C-0193 (Terence R. Mitchell and Lee Roy Beach, Investigators).